

Efficient computational strategies for Bayesian social networks

Alberto Caimo
Institute of Management
University of Lugano
Switzerland

Antonietta Mira
Institute of Finance
University of Lugano
Switzerland

March 19, 2014

Abstract

Powerful ideas recently appeared in the literature are adjusted and combined to design improved samplers for Bayesian exponential random graph models. Different forms of adaptive Metropolis-Hastings proposals (vertical, horizontal and rectangular) are tested and combined with the Delayed rejection (DR) strategy with the aim of reducing the variance of the resulting Markov chain Monte Carlo estimators for a given computational time. In the examples treated in this paper the best combination, namely horizontal adaptation with delayed rejection, leads to a variance reduction that varies between 92% and 144% relative to the adaptive direction sampling approximate exchange algorithm of Caimo and Friel (2011). These results correspond to an increased performance which varies from 10% to 94% if we take simulation time into account. The highest improvements are obtained when highly correlated posterior distributions are considered.

The methodology proposed in this paper has been implemented by the use of the `Bergm` package for R (Caimo and Friel, 2014).

1 Introduction

In this paper we combine the adaptive direction sampling approximate exchange algorithm (ADS-AEA) proposed in Caimo and Friel (2011), which has been proven to be particularly efficient in estimating exponential random graph models (ERGMs), with the delayed rejection (DR) introduced in Tierney and Mira (1999), a strategy to reduce the asymptotic variance of the resulting MCMC estimators. ADS-AEA is based on the idea of running in parallel multiple chains that, at each fixed simulation time, interact with each other to allow the construction of a proposal distribution that selects the proposal direction by picking at random a pair of chains.

We also suggest an alternative to ADS-AEA based on an adaptive random walk proposal distribution. Three different adaptation strategies will be studied to designing the proposal variance-covariance matrix: the first one is based on the past history of each single chain (vertical adaptation); the second is based on the current population of all chains at the given simulation time (horizontal adaptation), and finally global adaptation takes into account the past history of all chains (rectangular adaptation).

The three ingredients (ADS, DR and Adaptive proposal) are combined in various ways and compared to obtain the optimal strategy. Optimality is measure by the effective sample size (ESS) per simulation time and focus is on estimating ERGMs.

2 Exponential random graph models

Exponential random graph models (see Robins et al (2007) for a recent review) assume that the topological structure in an observed network y can be explained by the relative prevalence of a set of overlapping sub-graph configurations $s(y)$ also called graph or network statistics. Each configuration is assumed to have a particular probability of being observed in the given network: higher is the probability of being expressed in the graph, more are the chances of that statistic to occur and vice versa. The probability of a configuration being present in the network is expressed in terms of parameters. Configurations and parameters are at the core of ERGMs and, from a statistical point of view, the challenge is to estimate the parameters for each statistic such that the model is a good fit for the given data. From a statistical point of view, networks are relational data represented as mathematical graphs. A graph consists of a set of n nodes and a set of m ties which define some sort of relationships between pair of nodes called dyads. The connectivity pattern of a graph can be described by an $n \times n$ adjacency matrix y encoding the presence or absence of a tie between node i and j : $y_{ij} = 1$ if the dyad (i, j) is connected, $y_{ij} = 0$ otherwise. The likelihood of an ERGM represents the probability distribution of a random network graph and can be expressed as:

$$p(y|\theta) = \frac{\exp\{s(y)^T \theta\}}{z(\theta)}. \quad (1)$$

This equation states that the probability of observing a given network graph y is equal to the exponent of the observed graph statistics $s(y)$ multiplied by parameter vector θ divided by a normalising constant term $z(\theta)$. The latter is calculated over the sum of all possible graphs on n nodes and it is therefore extremely difficult to evaluate for all but trivially small graphs. The intractable normalising constant makes inference difficult for both frequentist and Bayesian approaches. This problem does not only occur in ERGMs, but in many other statistical models including, for example, the autologistic model (Besag, 1974) in spatial statistics. Given the similarities among these models from a computational tractability point of view, we envisage that the MCMC simulation strategies proposed in this paper are amenable of successful application in these other contexts as well.

3 Bayesian methods for ERGMs

Bayesian methods are becoming increasingly popular as techniques for modelling social networks. In the ERGM context recent works on using the Bayesian approach for inferring ERGMs have been proposed by Koskinen et al (2010) and Caimo and Friel (2011, 2013).

Following the Bayesian paradigm, prior distribution is assigned to θ . The posterior distribution of θ given the data y is:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}. \quad (2)$$

Direct evaluation of $p(\theta|y)$ requires the calculation of both the likelihood $p(y|\theta)$ and the marginal likelihood $p(y)$ which are typically intractable. For this reason posterior parameter estimation for ERGMs has been termed a doubly-intractable problem.

4 Exchange algorithm

Markov chain Monte Carlo (MCMC) algorithms (Tierney, 1994) are general simulation methods for sampling from posterior distributions and computing posterior quantities of interest. The

most widely used MCMC sampler is the Metropolis-Hastings (MH). A naïve MH update on the posterior distribution $p(\theta|y) \propto p(y|\theta)p(\theta)$ proposing to move from the current state θ to θ_1 , would require calculation of the following acceptance probability at each sweep of the algorithm:

$$\alpha(\theta, \theta_1) = 1 \wedge \frac{q(y|\theta_1)p(\theta_1)h(\theta|\theta_1)}{q(y|\theta)p(\theta)h(\theta_1|\theta)} \times \frac{z(\theta)}{z(\theta_1)} \quad (3)$$

where $q(\cdot)$ represents the unnormalised likelihood and $h(\cdot)$ is some proposal distribution. The ratio in (3) is unworkable due to the presence of the normalising constants $z(\theta)$ and $z(\theta_1)$.

Murray et al (2006) proposed to estimate $\frac{z(\theta)}{z(\theta_1)}$ directly, by considering the following augmented distribution:

$$p(\theta_1, y_1, \theta|y) \propto p(y|\theta)p(\theta)h(\theta_1|\theta) \times p(y_1|\theta_1) \quad (4)$$

where y_1 are auxiliary data generated by the distribution $p(\cdot|\theta_1)$ which is the same distribution from which the observed data y are assumed to have been sampled from. Notice that the original target is a proper marginal of the augmented distribution thus, running a Markov chain on the augmented state space and marginalizing over θ , returns an ergodic sample from the proper posterior of interest.

Using this augmented distribution has the advantage that the acceptance probability in (3) can be written as:

$$1 \wedge \frac{q(y|\theta_1)p(\theta_1)h(\theta|\theta_1)}{q(y|\theta)p(\theta)h(\theta_1|\theta)} \times \frac{q(y_1|\theta)}{q(y_1|\theta_1)} \times \frac{z(\theta)}{z(\theta_1)} \times \frac{z(\theta_1)}{z(\theta)}. \quad (5)$$

All intractable normalising constants cancel above and below the ratio making the acceptance probability (5) of the Metropolis-Hastings algorithm on the enlarged state space, computable.

5 Adaptive direction sampling approximate exchange algorithm (ADS-AEA)

The exchange algorithm of Murray et al (2006) requires exact simulation of new data y_1 from the likelihood $p(\cdot|\theta_1)$. However in the ERGM context, and more generally in Gibbs random fields, exact sampling from the likelihood is not possible. Caimo and Friel (2011) proposed to approximate the exact simulation of y_1 from $p(\cdot|\theta_1)$ using MCMC. A theoretical justification for the validity of this approach has been given by Everitt (2012).

In order to improve mixing Caimo and Friel (2011) use an adaptive direction sampling (ADS) method (Gilks et al, 1994; Roberts and Gilks, 1994) similar to that of ter Braak and Vrugt (2008). The approach consists of a collection of H chains which interact with one another. The ADS move, as illustrated in Caimo and Friel (2011), can be described as follows. Set a scalar value for γ (ADS move factor), for each chain h :

1. Sample two current states θ^{h_1} and θ^{h_2} without replacement from the population $\{1, \dots, H\} \setminus h$
2. Sample ϵ from a symmetric proposal distribution
3. Propose $\theta_1^h = \theta^h + \gamma (\theta^{h_1} - \theta^{h_2}) + \epsilon$
4. Accept the move from θ^h to θ_1^h with probability

$$\alpha(\theta^h, \theta_1^h) = 1 \wedge \frac{q(y_1|\theta^h)p(\theta_1^h)q(y|\theta_1^h)}{q(y|\theta^h)p(\theta^h)q(y_1|\theta_1^h)}. \quad (6)$$

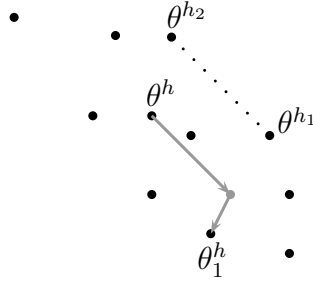


Figure 1: The move of θ^h is generated from the difference $\theta^{h_1} - \theta^{h_2}$ plus a random term ϵ .

Note that, since the ADS proposal distribution is symmetric, it does not appear in the acceptance probability.

5.1 Florentine marriage network

Let us consider, as a toy example, the 16-node Florentine marriage network data concerning the marriage relations between some Florentine families in around 1430 (Padgett and Ansell, 1993). The network graph is displayed in Figure 2.

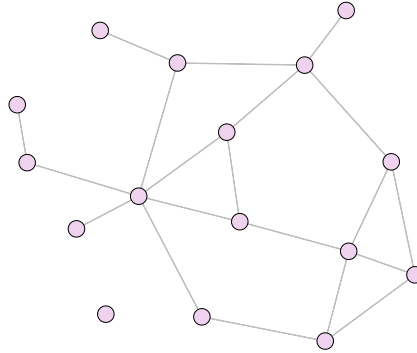


Figure 2: Florentine marriage network graph.

We propose to estimate the posterior distribution of the following 3-dimensional ERGM:

$$p(y|\theta) \propto \exp \left\{ \theta^{(1)} s_1(y) + \theta^{(2)} s_2(y) + \theta^{(3)} s_3(y) \right\} \quad (7)$$

where

$$\begin{aligned} s_1(y) &= \sum_{i < j} y_{ij} && \text{number of edges} \\ s_2(y) &= \sum_{i < j < k} y_{ik} y_{jk} && \text{number of 2-stars} \\ s_3(y) &= \sum_{i < j < l < k} y_{ik} y_{jk} y_{lk} && \text{number of 3-stars.} \end{aligned}$$

A flat multivariate Normal prior $p(\theta) \sim \mathcal{N}(0, 100I_d)$ is chosen, where I_d is the identity matrix with dimensions equal to that of the model (the same prior setting will be used for all the examples in this paper). The **Bergm** package for R (Caimo and Friel, 2014) allows to carry out inference with the approximate exchange algorithm described above.

We set the ADS move factor $\gamma = 0.8$ and $\epsilon \sim \mathcal{N}(0, 0.025I_d)$. The auxiliary chain consists of 50 iterations and the main chain of 4,000 iterations for each of the 6 chains of the MCMC

Table 1: Florentine marriage network - Posterior parameter estimates and effective sample size (ESS).

| | $\theta^{(1)}$ (edges) | $\theta^{(2)}$ (2-stars) | $\theta^{(3)}$ (3-stars) |
|------------|------------------------|--------------------------|--------------------------|
| Post. mean | -1.57 | 0.08 | -0.07 |
| Post. sd | 1.93 | 0.71 | 0.34 |
| ESS | 736 | 743 | 760 |

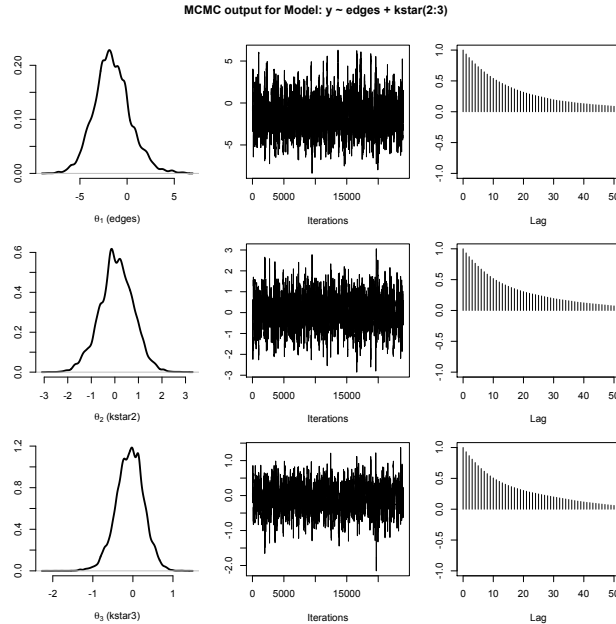


Figure 3: Florentine marriage network. MCMC diagnostics for the overall chain. The 3 plot columns, from left to right, are: estimated marginal posterior densities, trace plots, and auto-correlation plots.

population so that we have a total of 24,000 main iterations. The tuning parameters were chosen so that the overall acceptance rate is around 21%. Table 1 shows the posterior estimates and effective sample size (ESS) (Kass et al, 1998).

The results indicate the tendency to a low number of edges as expressed by the edge parameter $\theta^{(1)}$ and null parameter values for $\theta^{(2)}$ and $\theta^{(3)}$. These estimates are consistent with the ones obtained using a frequentist approach (Hunter et al, 2008) as expected given the fairly flat prior.

6 Delayed rejection strategy

Delayed rejection (DR) is a modification of the Metropolis-Hastings MCMC algorithm introduced in Tierney and Mira (1999) and generalized in Green and Mira (2001); Mira (2001a), aimed at improving efficiency of the resulting MCMC estimators relative to asymptotic variance orderings introduced in Peskun (1973) and generalized by Tierney (1998); Mira (2001b). The basic idea is that, upon rejection in a MH, instead of advancing time and retaining the

same position, a second stage move is proposed. The acceptance probability of the second stage candidate preserves reversibility of the Markov chain with respect to the target distribution of interest (the posterior, in a Bayesian setting). This delaying rejection mechanism can be iterated for a fixed or random number of stages. The higher stage proposal distributions are allowed to depend on the candidates so far proposed and rejected. Thus DR allows partial local adaptation of the proposal within each time step of the Markov chain still retaining reversibility and the Markovian property. The advantage of DR over alternative ways of combining different MH proposals or kernels, such as mixing and cycling (Tierney, 1994), is that a hierarchy between kernels can be exploited so that kernels that are easier to compute (in terms of CPU time) are tried first, thus saving in terms of simulation time. Or moves that are more “bold” (bigger variance of the proposal, for example) are tried at earlier stages thus allowing the sampler to explore the state space more efficiently following a sort of “first bold” versus “second timid” tennis-service strategy. To simplify the notation indicate the posterior distribution of interest, $p(\theta|y)$, simply as $p(\theta)$, dropping the conditioning on y . Suppose the current position of the Markov chain is $X_t = \theta$. As in a regular MH, a candidate move θ_1 is generated from a proposal $h_1(\theta, \cdot)$ and accepted with probability

$$\alpha_1(\theta, \theta_1) = 1 \wedge \frac{p(\theta_1)h_1(\theta_1|\theta)}{p(\theta)h_1(\theta|\theta_1)} = 1 \wedge \frac{N_1}{D_1}. \quad (8)$$

Note that the subscript in h_1 and α_1 indicate that this is the first stage proposal and acceptance probability. Upon rejection, instead of retaining the same position, $X_{t+1} = \theta$, as we would do in a standard MH, a second stage move θ_2 is generated from a proposal distribution that is allowed to depend, not only on the current position of the chain, but also on what we have just proposed and rejected: $h_2(\theta_2|\theta, \theta_1)$. The second stage acceptance probability is:

$$\begin{aligned} \alpha_2(\theta, \theta_1, \theta_2) &= 1 \wedge \frac{p(\theta_2)h_1(\theta_1|\theta_2)h_2(\theta|\theta_2, \theta_1)[1 - \alpha_1(\theta_2, \theta_1)]}{p(\theta)h_1(\theta_1|\theta)h_2(\theta_2|\theta, \theta_1)[1 - \alpha_1(\theta, \theta_1)]} \\ &= 1 \wedge \frac{N_2}{D_2}. \end{aligned} \quad (9)$$

This process of delaying rejection can be iterated and the i -th stage acceptance probability is, following Mira (2001a):

$$\begin{aligned} \alpha_i(\theta, \theta_1 \dots \theta_i) &= \\ &= 1 \wedge \frac{N_i}{D_i} \\ &= 1 \wedge \left\{ \frac{p(\theta_i)h_1(\theta_{i-1}|\theta_i)h_2(\theta_{i-2}|\theta_i, \theta_{i-1}) \dots h_i(\theta|\theta_i, \theta_{i-1} \dots \theta_1)}{p(\theta)h_1(\theta_1|\theta)h_2(\theta_2|\theta, \theta_1) \dots h_i(\theta_i|\theta, \theta_1 \dots \theta_{i-1})} \right. \\ &\quad \left. \frac{[1 - \alpha_1(\theta_i, \theta_{i-1})][1 - \alpha_2(\theta_i, \theta_{i-1}, \theta_{i-2})] \dots [1 - \alpha_{i-1}(\theta_i, \dots, \theta_1)]}{[1 - \alpha_1(\theta, \theta_1)][1 - \alpha_2(\theta, \theta_1, \theta_2)] \dots [1 - \alpha_{i-1}(\theta, \theta_1, \dots, \theta_{i-1})]} \right\} \end{aligned} \quad (10)$$

If the i -th stage is reached, it means that $N_j < D_j$ for $j = 1, \dots, i-1$, therefore $\alpha_j(\theta, \theta_1 \dots \theta_j)$ is simply N_j/D_j , $j = 1, \dots, i-1$ and a recursive formula can be obtained: $D_i = h_i(\theta \dots \theta_i)(D_{i-1} - N_{i-1})$ which leads to:

$$\begin{aligned} D_i &= h_i(\theta_i|\theta \dots) [h_{i-1}(\theta_{i-1}|\theta \dots) [h_{i-2}(\theta_{i-2}|\theta \dots) \dots \\ &\quad [h_2(\theta_2, |\theta, \theta_1) [h_1(\theta_1|\theta)p(\theta) - N_1] - N_2] - N_3] \dots \\ &\quad - N_{i-1}]. \end{aligned} \quad (11)$$

Since reversibility with respect to p is preserved separately at each stage, the process of delaying rejection can be interrupted at any stage. The user can either decide, in advance, to try at most, a fixed number of times to move away from the current position or, alternatively, upon each rejection, toss a p -coin (i.e. a coin with head probability equal to p), and if the outcome is head move to a higher stage proposal, otherwise stay put.

Tierney and Mira (1999) prove that the DR strategy provides MCMC estimators with smaller asymptotic variance than standard MH. This better performance holds no matter what is the function f whose expectation relative to the target posterior we want to estimate (provided f is squared integrable with respect to the target).

7 Approximate Exchange Algorithm with Delayed Rejection (AEA+DR)

The idea is to combine the DR strategy with the Approximate Exchange Algorithm. We name this new algorithm the AEA+DR and different instances of it will be specified in subsequent sections depending of the (adaptive) proposal distribution used. For the AEA+DR algorithm a theoretical modification of the i -th stage acceptance probability is needed to take into account the fact that the target normalising constant depends on the parameter of interest. This is a novel methodological contribution that gives rise to an efficient MCMC sampler that can be used in general for doubly intractable problems.

The first stage acceptance probability is unchanged relative to the standard AEA, and (re-calling (5)) is given by:

$$\alpha_1(\theta, \theta_1) = 1 \wedge \frac{q(y_1|\theta) p(\theta_1) h_1(\theta|\theta_1) q(y|\theta_1)}{q(y|\theta) p(\theta) h_1(\theta_1|\theta) q(y_1|\theta_1)} \quad (12)$$

The second stage acceptance probability that preserves the detailed balance condition is:

$$\alpha_2(\theta, \theta_1, \theta_2) = 1 \wedge \frac{q(y_2|\theta) p(\theta_2) h_1(\theta_1|\theta_2) h_2(\theta|\theta_2, \theta_1) q(y|\theta_2) [1 - \alpha_1(\theta_2, \theta_1)]}{q(y|\theta) p(\theta) h_1(\theta_1|\theta) h_2(\theta_2|\theta, \theta_1) q(y_2|\theta_2) [1 - \alpha_1(\theta, \theta_1)]} \quad (13)$$

Higher stage acceptance probabilities are modified accordingly.

The second stage proposal, h_2 , can be designed in a very flexible way (using our intuition on the target at hand) and is allowed to depend not only on the current position of the Markov chain (θ), but also the previously rejected proposal (θ_1). In some sense we can learn from our earlier mistakes. A similar reasoning holds true for higher stage proposals. The second stage proposal of the delayed rejection version of the adaptive direction sampler (named ADS+DR) is designed to be negatively correlated with the first stage one following the idea of antithetic second stage suggested in Bédard et al (2010).

8 Adaptive Approximate Exchange Algorithm (AAEA)

Three forms of adaptation of the Metropolis-Hastings proposal distribution (alternative to the AEA) are considered: vertical, horizontal and rectangular. At simulation time t there is a rectangular $t \times H$ family of particles available: $\theta_{i,j}, i = 1, \dots, t; j = 1, \dots, H$. Suppose we

are interested in updating the position of particle $\theta_{t,h}$ (in the previous formulas this particle was simply indicated as θ with no subscripts). To this aim a random walk Metropolis-Hastings proposal is designed by taking a Gaussian distribution with mean equal to $\theta_{t,h}$ and variance-covariance matrix given by the empirical variance of either:

- AAEA-1 all past particles along the same chain h (vertical adaptation): $\theta_{i,h}, i = 1, \dots, t-1$;
- AAEA-2 all particles at the current time t for all chains (horizontal adaptation): $\theta_{t,i}, i = 1, \dots, H$;
- AAEA-3 particles from all chains and all past simulations (rectangular adaptation, aka inter-chain adaptation from Craiu et al, 2009): $\theta_{i,j}, i = 1, \dots, t-1; j = 1, \dots, H$.

In order to meet the conditions of Roberts and Rosenthal (2007) that guarantee that, despite adaptation, the proper stationary distribution is preserved and ergodicity holds, with a small probability β (set equal to 0.01 in our simulation study), instead of using the adaptive proposal described above the following static proposal is used: Normal distribution with variance-covariance matrix equal to $0.0025I_d$ where I_d is the identity matrix of size d , where d is the model dimension. As also discussed in Craiu et al (2009), a question of interest in adaptive MCMC is whether one should wait a short or a long time before starting the adaptation. Based on our simulation experience we found that the most effective strategy is to use the ADS approach during the burn-in phase and then switch to one of the adaptive algorithm mentioned above. Furthermore, the simulation results presented use intensive adaptation (Giordani and Kohn, 2010) i.e. adaptation is performed at every iteration.

We believe that more sophisticated forms of adaptations such as regional or tempered adaptation (Craiu et al, 2009) are not needed in our setting.

9 Adaptive Approximate Exchange Algorithm with Delayed Rejection (AAEA+DR)

The three adaptation schemes outlined in the previous section could be combined within the DR mechanism. For example at first stage horizontal adaptation can be used since the resulting proposal is typically less computationally intensive to obtain (because H is usually smaller than t especially after the burn-in phase), at second stage we can try vertical adaptation and resort to rectangular adaptation only at third stage. The intuition behind this combination is to use simple proposals first and resort to more refined proposals (typically more computationally intensive to construct, as rectangular adaptation) only if really needed.

In the examples considered we follow a different rationale when combining the Adaptive Approximate Exchange Algorithm with the delayed rejection, namely, the second stage proposal is equal to the first stage one with the variance-covariance matrix rescaled by a factor of 0.5. In other words a more timid move is attempted at second stage. This is a very naïve form of delayed rejection but it is often quite effective.

10 Examples

In this section we compare the adaptive direction sampler (ADS-AEA) with three alternative forms of adaptation as defined in Section 8. The comparison takes into account the effective sample size (ESS) and the performance (i.e. ESS divided by simulation time).

For each one of the four algorithms considered we have also implemented the corresponding two stage DR version as explained in Section 7. For vertical, horizontal and rectangular adaptation this is done by simply adding a second stage proposal which is identical to the first stage one except that the variance-covariance matrix is multiplied by 0.5.

For the DR version of the adaptive direction sampler, upon rejection of the first stage candidate move $\theta_1^h = \theta^h + \gamma_1 (\theta_1^{h1} - \theta_1^{h2}) + \epsilon_1$, the second stage proposal is deterministically obtained from the first one by simply going in the opposite direction relative to the current position: $\theta_2^h = \theta^h - \gamma_1 (\theta_1^{h1} - \theta_1^{h2}) + \epsilon_1$ or, in other terms, $\theta_2^h = 2\theta_0^h - \theta_1^h$. This strategy follows the idea of second stage antithetic proposal which has been proved in (Bédard et al, 2010) to be highly effective.

We thus have a total of 8 different algorithms under comparison. To our surprise the horizontal adaptation approach AAEA-2 outperforms all other adaptive algorithms (we were expecting rectangular adaptation to have a better performance given that more particles are used to learn the variance-covariance structure of the target distribution that is then used in the random walk Metropolis-Hastings sampler). The performance of the horizontal adaptive algorithms is then further enhanced when combined with the delayed rejection strategy. The improvements of this approach over the competitors are more evident in the case of models with highly correlated posterior densities such as Examples 10.1 and 10.3.

The DR version of each algorithm is always better than the corresponding single stage proposal in terms of ESS. This simply confirms the fact that the DR dominates the corresponding single stage Metropolis-Hastings sampler in the Peskun ordering i.e. in terms of asymptotic variance of the resulting MCMC estimators (for a given number of sweeps). If the additional simulation time of the DR is also taken into account in the comparison (i.e. if we consider the ESS per unit computational time), the DR still outperforms the corresponding single stage proposal in all cases but for the adaptive direction sampler (this is because the second stage proposal of the adaptive direction sampler, ADS+DR, has a very small acceptance probability due to the structure of the first stage proposal and the antithetic move implemented at second stage).

10.1 Florentine marriage network

Let us consider again the Florentine marriage network and model defined in Equation 7. We use a total number of 24,000 iterations for estimating the posterior density and 4 different approaches: the ADS-AEA, AAEA-1 (vertical adaptation) and AAEA-3 (rectangular adaptation) consisting of 6 chains of 4,000 iterations each, and AAEA-2 (horizontal adaptation) consisting of 24 chains of 1,000 iterations each.

In Table 2 are displayed the posterior parameter estimates and effective sample size calculated for the AAEA-2 and AAEA-2+DR which turned out to be the best approaches in terms of performance. In particular the AAEA-2+DR yields a variance reduction of about 63% compared to the ADS-AEA.

In Figure 5 it can be seen that the autocorrelations of the parameter estimates returned by the AAEA-2+DR decay quicker than the autocorrelations of the other two approaches displayed in Figures 3 and 4. The AAEA-2 algorithm outperforms the ADS-AEA in terms of both ESS (24%) and performance (41%). The AAEA-2+DR outperforms the AAEA-2 in terms of ESS of about 97% and performance of about 16%.

In Table 3 it is possible to observe the correlation matrix between the parameters in the posterior distribution. There is a very strong negative correlation between all the parameters of the model.

Table 2: Florentine marriage network - Posterior parameter estimates and effective sample size (ESS) for model 7.

| | AAEA-2 (horizontal adaptation) | | |
|------------|--|--------------------------|--------------------------|
| | $\theta^{(1)}$ (edges) | $\theta^{(2)}$ (2-stars) | $\theta^{(3)}$ (3-stars) |
| Post. mean | -1.47 | 0.05 | - 0.06 |
| Post. sd | 1.86 | 0.69 | 0.36 |
| ESS | 894 | 948 | 928 |
| | AAEA-2+DR (horizontal adaptation + DR) | | |
| | $\theta^{(1)}$ (edges) | $\theta^{(2)}$ (2-stars) | $\theta^{(3)}$ (3-stars) |
| Post. mean | -1.61 | 0.08 | - 0.06 |
| Post. sd | 1.55 | 0.53 | 0.25 |
| ESS | 1712 | 1913 | 1850 |

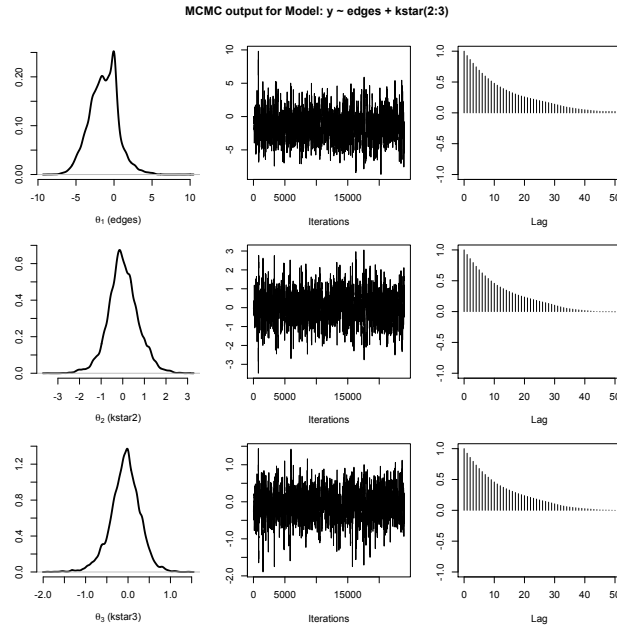


Figure 4: Florentine marriage network - MCMC diagnostics for the AAEA-2.

Table 3: Florentine marriage network - Posterior correlation matrix between the parameters in the distribution for model 7.

| | $\theta^{(1)}$ | $\theta^{(2)}$ | $\theta^{(3)}$ |
|----------------|----------------|----------------|----------------|
| $\theta^{(1)}$ | 1.00 | -0.94 | -0.80 |
| $\theta^{(2)}$ | . | 1.00 | -0.94 |
| $\theta^{(3)}$ | . | . | 1.00 |

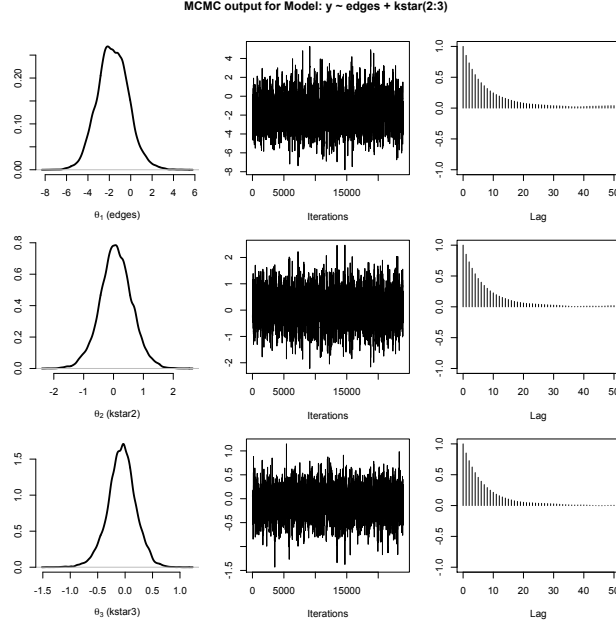


Figure 5: Florentine marriage network - MCMC diagnostics for the AAEA-2+DR.

10.2 Teenage Friends and Lifestyle Study

In this example we revisit a well known study in social science concerning adolescent friendship relations in a secondary school in Glasgow, Scotland (Pearson and Michell, 2000). Sport activity of each of the 50 students is considered as a covariate information x (see Figure 6).

The main focus is on the transitivity effect expressed by the number of triangles observed among actors with the same covariate and the differential homophily effect between students practicing sport with the same regularity: group 1 indicates students with non-regular sport activity and group 2 indicates students with regular sport activity. The model can be written as:

$$p(y|\theta) \propto \exp \left\{ \theta^{(1)} s_1(y) + \theta^{(2)} s_2(y, x) + \theta^{(3)} s_3(y, x) + \theta^{(4)} s_4(y, x) \right\} \quad (14)$$

where:

$$\begin{aligned} s_1(y) &= \sum_{i < j} y_{ij} && \text{number of edges} \\ s_2(y, x) &= \sum_{i < j} y_{ij} \mathbb{1}_{(x_i = x_j)} && x = \text{group 1} \\ s_3(y, x) &= \sum_{i < j} y_{ij} \mathbb{1}_{(x_i = x_j)} && x = \text{group 2} \\ s_4(y, x) &= \sum_{i < j < k} y_{ij} y_{ik} y_{jk} \mathbb{1}_{(x_i = x_j = x_k)} && \text{triangles} \end{aligned}$$

and $\mathbb{1}(\cdot)$ is the indicator function.

The tuning parameters for the ADS proposal $\gamma = 0.7$ and $\epsilon \sim \mathcal{N}(0, 0.0025I_d)$ are chosen so as to obtain the overall acceptance rate is around 21%. 200 auxiliary iterations are used for network simulation and 32,000 main iterations are used for estimating the posterior density of model defined in Equation 14: the ADS-AEA consists of 8 chains, AAEA-2 consists of 32 chains, and AAEA-2+DR consists of 16 chains.

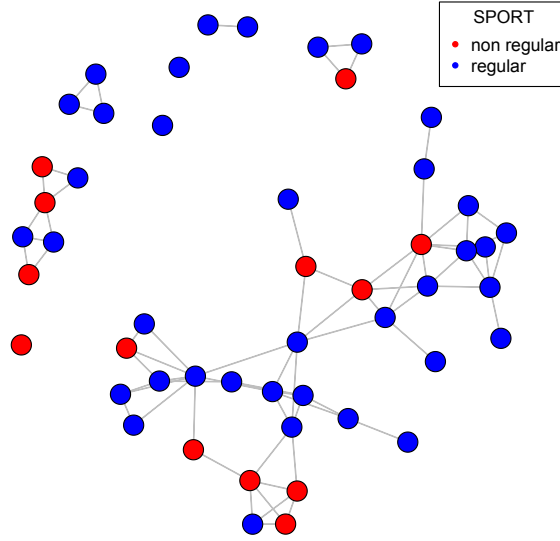


Figure 6: Teenage friendship network graph.

In this case, both the AAEEA-1 and AAEEA-3 outperforms the AAEEA-2 in terms of variance reduction but not in terms of performance. So the AAEEA-2 is still to be preferred.

In Table 5 it is possible to observe the correlation matrix between the parameters in the posterior distribution: the strongest (negative) correlation is between parameters $\theta^{(1)}$ associated to the number of edges and $\theta^{(3)}$ associated to the homophily effect within group 2.

In Figure 9 it can be seen that the autocorrelations of the parameters estimated by the AAEEA-2+DR approach decay quicker than the autocorrelations estimated by the other approaches (Figure 7 and Figure 8). The AAEEA-2 and the ADS-AAEA are equivalent in terms of both variance reduction and performance. Delayed rejection turned out to be particularly efficient when combined with the AAEEA-2. In fact in terms of variance reduction the AAEEA-2+DR is about 90% better than the ADS-AAEA and about 10% in terms of performance.

From the results displayed in Table 4 we can conclude that teenagers having the same level of sport activity do not seem to create a significant number of friendship connections. The transitivity effect within each group effect expressed by $\theta^{(4)}$ is an important feature of the network.

10.3 Karate club network

This example concerns the karate club network (Zachary, 1977) displayed in Figure 10 which represents friendship relations between 34 members of a karate club at a US university in the 1970. We propose to estimate the following 3-dimensional model using the network statistics proposed by Snijders et al (2006):

$$p(y|\theta) \propto \exp \left\{ \theta^{(1)} s_1(y) + \theta^{(2)} v(y, \phi_u) + \theta^{(3)} u(y, \phi_v) \right\} \quad (15)$$

where

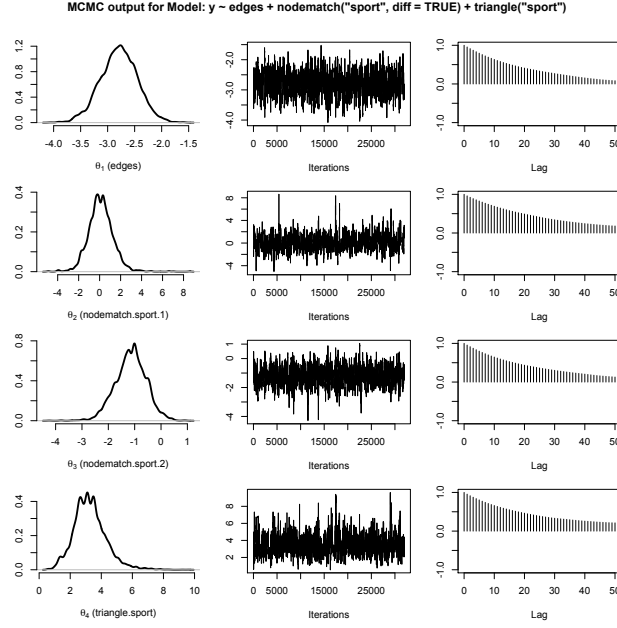


Figure 7: Teenage friendship network - MCMC diagnostics for the ADS-AEA.

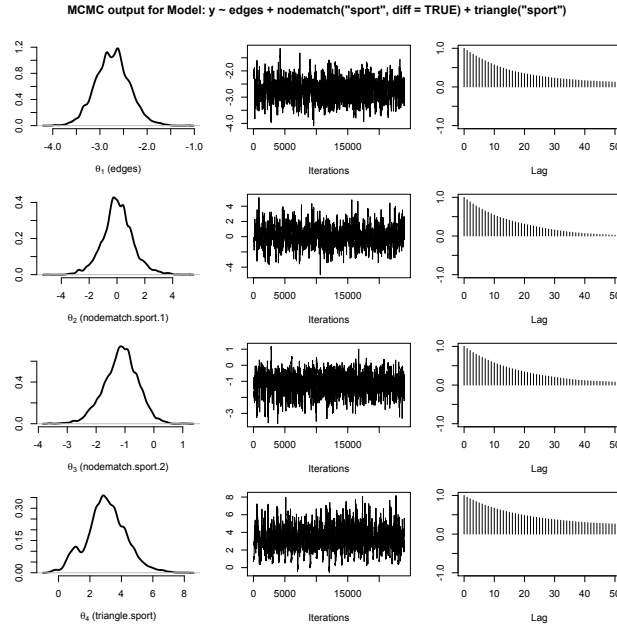


Figure 8: Teenage friendship network - MCMC diagnostics for the AAEA-2.

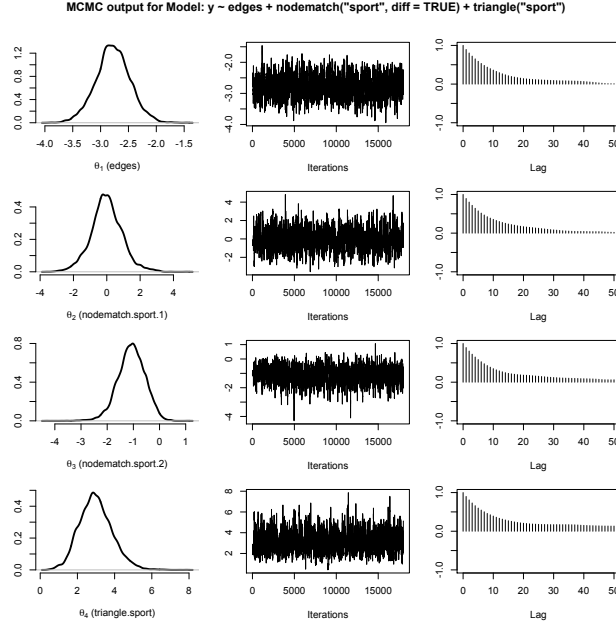


Figure 9: Teenage friendship network - MCMC diagnostics for the AAEA-2+DR.

Table 4: Teenage friendship network - Posterior parameter estimates and effective sample size (ESS) for model 14.

| | ADS-AEA | | | |
|------------|--|-----------------------------------|-----------------------------------|----------------------------------|
| | $\theta^{(1)}$ (edges) | $\theta^{(2)}$ (homophily.group1) | $\theta^{(3)}$ (homophily.group2) | $\theta^{(4)}$ (triangle.groups) |
| Post. mean | -2.78 | -0.02 | -1.01 | 2.73 |
| Post. sd | 0.26 | 0.80 | 0.45 | 0.70 |
| ESS | 829 | 900 | 827 | 887 |
| | AAEA-2 (horizontal adaptation) | | | |
| | $\theta^{(1)}$ (edges) | $\theta^{(2)}$ (homophily.group1) | $\theta^{(3)}$ (homophily.group2) | $\theta^{(4)}$ (triangle.groups) |
| Post. mean | -2.78 | -0.01 | -1.03 | 2.79 |
| Post. sd | 0.26 | 0.79 | 0.46 | 0.75 |
| ESS | 822 | 953 | 825 | 651 |
| | AAEA-2+DR (horizontal adaptation + DR) | | | |
| | $\theta^{(1)}$ (edges) | $\theta^{(2)}$ (homophily.group1) | $\theta^{(3)}$ (homophily.group2) | $\theta^{(4)}$ (triangle.groups) |
| Post. mean | -2.77 | -0.03 | -0.95 | 2.58 |
| Post. sd | 0.27 | 0.75 | 0.46 | 0.70 |
| ESS | 1690 | 1749 | 1655 | 1516 |

Table 5: Teenage friendship network - Posterior correlation matrix between the parameters in the distribution for model 14.

| | $\theta^{(1)}$ | $\theta^{(2)}$ | $\theta^{(3)}$ | $\theta^{(4)}$ |
|----------------|----------------|----------------|----------------|----------------|
| $\theta^{(1)}$ | 1.00 | -0.37 | -0.54 | -0.02 |
| $\theta^{(2)}$ | . | 1.00 | 0.28 | 0.15 |
| $\theta^{(3)}$ | . | . | 1.00 | -0.47 |
| $\theta^{(4)}$ | . | . | . | 1.00 |

$s_1(y) = \sum_{i < j} y_{ij}$ number of edges

$v(y, \phi_v) = e^{\phi_v} \sum_{i=1}^{n-2} \left\{ 1 - (1 - e^{-\phi_v})^i \right\} EP_i(y)$ geometrically weighted edgewise shared partners (GWESP)

$u(y, \phi_u) = e^{\phi_u} \sum_{i=1}^{n-1} \left\{ 1 - (1 - e^{-\phi_u})^i \right\} D_i(y)$ geometrically weighted degrees (GWD)

where $EP_i(y)$ and $D_i(y)$ are the edgewise shared partners and degree distributions respectively. We set $\phi_u = \phi_v = \log(2)$ so that the model is a non-curved ERGM (Hunter and Handcock, 2006). The prior setting is analogous to the one in Section 5.1: $p(\theta) \sim \mathcal{N}(0, 100I_3)$. The tuning parameters for the ADS proposal are: $\gamma = 0.5$ and $\epsilon \sim \mathcal{N}(0, 0.0025I_d)$ so that the overall acceptance rate is around 21%. The auxiliary chain consists of 100 iterations and a total number of 24,000 main iterations is used. The number of chains used in the various strategies is analogous to the previous example in Section 10.1.

In this example, as happened in the teenage friendship network example above, the AAEA-3 outperforms the AAEA-2 in terms of variance reduction of about 40% but not in terms of performance. For this reason AAEA-2 is still to be preferred.

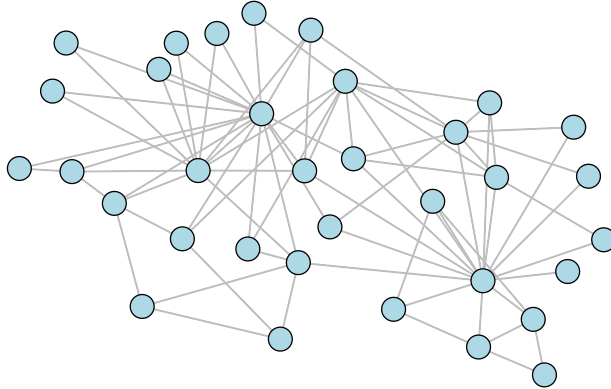


Figure 10: Zachary karate club network graph.

In Figure 13 it can be seen that the autocorrelations of the parameters for the AAEA-2 approach decay quicker than the autocorrelations given by the other methods as shown in Figure 11 and 12. The AAEA-2 outperforms the ADS-AEA of about 12% in terms of performance whereas the AAEA-2+DR makes a further improvement of about 70% (see Table 6).

As in the Florentine marriage network example, we can observe (Table 7) that there is a strong negative posterior correlation between parameters $\theta^{(1)}$ and $\theta^{(2)}$ and between $\theta^{(1)}$ and $\theta^{(3)}$.

Generally a strong correlation between parameters in the posterior distribution hampers the behaviour of the MCMC scheme. In fact high posterior correlation can slow down the motion

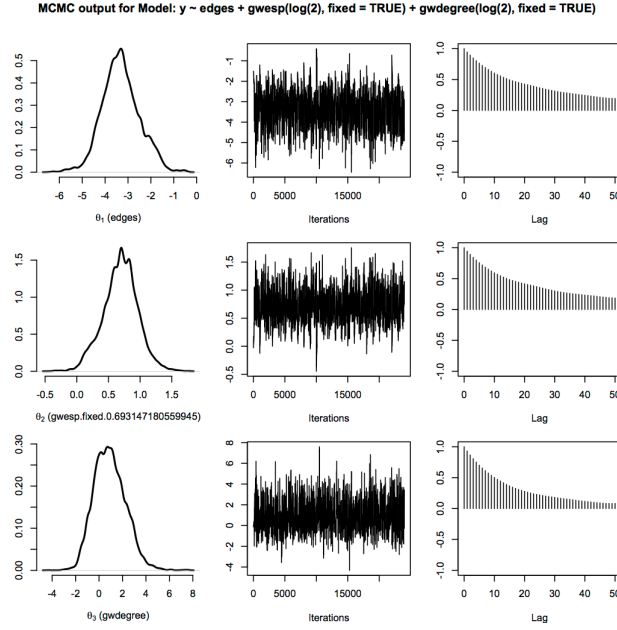


Figure 11: Zachary karate club network - MCMC diagnostics for the ADS-AEA.

Table 6: Zachary karate club network - Posterior parameter estimates and effective sample size (ESS) for model 15.

| | ADS-AEA | | |
|------------|--|------------------------|---------------------------|
| | $\theta^{(1)}$ (edges) | $\theta^{(2)}$ (gwesp) | $\theta^{(3)}$ (gwdegree) |
| Post. mean | -3.51 | 0.74 | 1.18 |
| Post. sd | 0.62 | 0.21 | 1.12 |
| ESS | 760 | 703 | 805 |
| | AAEA-2 (horizontal adaptation) | | |
| | $\theta^{(1)}$ (edges) | $\theta^{(2)}$ (gwesp) | $\theta^{(3)}$ (gwdegree) |
| Post. mean | -3.30 | 0.68 | 0.99 |
| Post. sd | 0.73 | 0.27 | 1.22 |
| ESS | 479 | 493 | 837 |
| | AAEA-2+DR (horizontal adaptation + DR) | | |
| | $\theta^{(1)}$ (edges) | $\theta^{(2)}$ (gwesp) | $\theta^{(3)}$ (gwdegree) |
| Post. mean | -3.44 | 0.72 | 1.01 |
| Post. sd | 0.59 | 0.21 | 1.07 |
| ESS | 1647 | 1723 | 1521 |

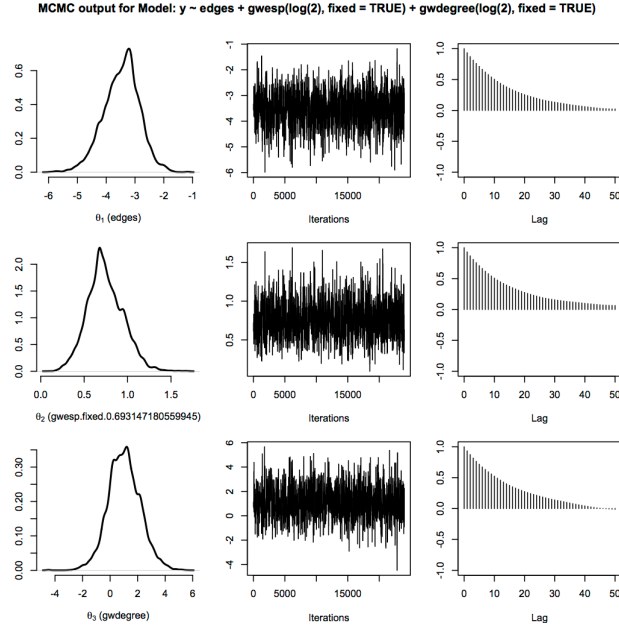


Figure 12: Zachary karate club network - MCMC diagnostics for the AAEA-2.

Table 7: Zachary karate club network - Posterior correlation matrix between the parameters in the distribution for model 15.

| | $\theta^{(1)}$ | $\theta^{(2)}$ | $\theta^{(3)}$ |
|----------------|----------------|----------------|----------------|
| $\theta^{(1)}$ | 1.00 | -0.80 | -0.75 |
| $\theta^{(2)}$ | . | 1.00 | 0.37 |
| $\theta^{(3)}$ | . | . | 1.00 |

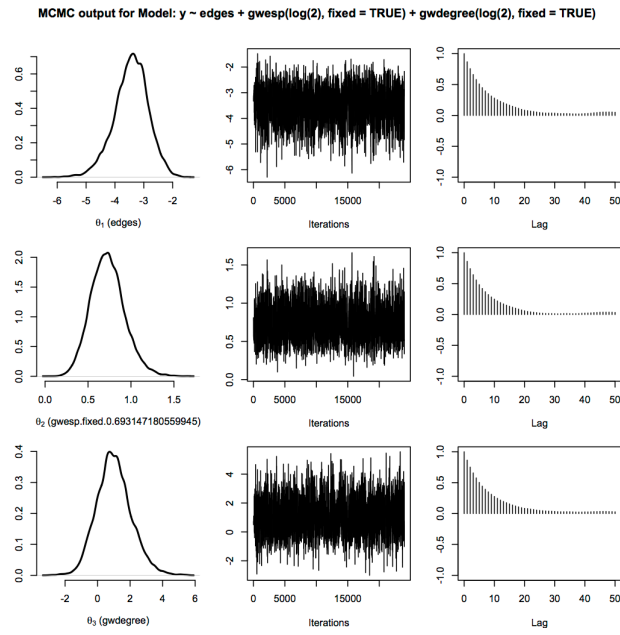


Figure 13: Zachary karate club network - MCMC diagnostics for the AAEA-2+DR.

of the chain towards equilibrium distribution. It is in this case that the adaptive approximate exchange algorithm with delayed rejection (AAEA-2+DR) gives the best performance compared to the adaptive direction sampling approximate exchange algorithm.

11 Conclusions

The exchange algorithm of Murray et al (2006) makes the computation of the MH acceptance probability feasible even for target distributions whose normalizing constant depends on the parameter of interest (doubly intractable problems). The *approximate exchange algorithm*, due to Caimo and Friel (2011), modifies the original exchange algorithm and makes it applicable also in settings where sampling from the assumed data generating process is not feasible. This is the case for exponential random graphs the model we focus on in this paper.

The *delayed rejection* strategy allows to locally adapt the proposal distribution within each sweep of a MH algorithm at the cost of additional computational time.

The *adaptive random walk proposal* of Haario et al (2001) revised by Roberts and Rosenthal (2009) allows for global adaptation between MH iterations. This learning from the past process is also expensive from a computational point of view.

These three ingredients are combined in different ways to achieve variance reduction relative to the adaptive direction sampling approximate exchange algorithm of Caimo and Friel (2011) implemented in the *Bergm* package for R (Caimo and Friel, 2014), which is our benchmark.

The 8 algorithms under comparison (seven of which are original contributions) are tested on three examples. Consistently, the best combination (in terms of ESS for fixed simulation time), is given by the *horizontal adaptive approximate exchange algorithm with delayed rejection*, which achieves a variance reduction that varies between 92% and 144% (relative to the benchmark). This translates into a better performance varying from 10% to 94%, if the extra simulation time, due to the delayed rejection mechanism and the adaptation procedure, is taken into account. The strongest improvements are obtained in the examples with highly correlated posterior distributions.

The delayed rejection strategy and the form of adaptation proposed in the present paper are under implementation in the *Bergm* package.

References

- Bédard M, Douc R, Moulines E (2010) Scaling analysis of delayed rejection MCMC methods. *Methodology and Computing in Applied Probability* pp 1–28
- Besag JE (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B* 36:192–236
- ter Braak CJ, Vrugt JA (2008) Differential evolution Markov chain with snooker updater and fewer chains. *Statistics and Computing* 18(4):435–446
- Caimo A, Friel N (2011) Bayesian inference for exponential random graph models. *Social Networks* 33(1):41 – 55
- Caimo A, Friel N (2013) Bayesian model selection for exponential random graph models. *Social Networks* 35(1):11 – 24

- Caimo A, Friel N (2014) Bergm: Bayesian exponential random graphs in R. *Journal of Statistical Software* (to appear)
- Craiu RV, Rosenthal J, Yang C (2009) Learn from thy neighbor: Parallel-chain and regional adaptive MCMC. *Journal of the American Statistical Association* 104(488):1454–1466
- Everitt RG (2012) Bayesian parameter estimation for latent Markov random fields and social networks. *Journal of Computational and Graphical Statistics* 21(4):940–960
- Gilks WR, Roberts GO, George EI (1994) Adaptive direction sampling. *Statistician* 43(1):179–189
- Giordani P, Kohn R (2010) Adaptive independent Metropolis–Hastings by fast estimation of mixtures of normals. *Journal of Computational and Graphical Statistics* 19(2):243–259
- Green PJ, Mira A (2001) Delayed rejection in reversible jump Metropolis–Hastings. *Biometrika* 88(4):1035–1053
- Haario H, Saksman E, Tamminen J (2001) An adaptive Metropolis algorithm. *Bernoulli* 7(2):223–242
- Hunter DR, Handcock MS (2006) Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics* 15:565–583
- Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M (2008) ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software* 24(3):1–29, URL <http://www.jstatsoft.org/v24/i03>
- Kass RE, Carlin BP, Gelman A, Neal RM (1998) Markov chain Monte Carlo in practice: A roundtable discussion. *The American Statistician* 52(2):93–100
- Koskinen JH, Robins GL, Pattison PE (2010) Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation. *Statistical Methodology* 7(3):366–384
- Mira A (2001a) On Metropolis-Hastings algorithms with delayed rejection. *Metron* 59(3-4):231–241
- Mira A (2001b) Ordering and improving the performance of Monte Carlo Markov chains. *Statistical Science* pp 340–350
- Murray I, Ghahramani Z, MacKay D (2006) MCMC for doubly-intractable distributions. In: *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, AUAI Press, Arlington, Virginia
- Padgett JF, Ansell CK (1993) Robust action and the rise of the Medici, 1400–1434. *American Journal of Sociology* 98:1259–1319
- Pearson M, Michell L (2000) Smoke rings: social network analysis of friendship groups, smoking and drug-taking. *Drugs: Education, Prevention and Policy* 7(1):21–37
- Peskun P (1973) Optimum monte-carlo sampling using Markov chains. *Biometrika* 60(3):607–612
- Roberts GO, Gilks WR (1994) Convergence of adaptive direction sampling. *Journal of Multivariate Analysis* 49(2):287–298

- Roberts GO, Rosenthal JS (2007) Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *Journal of applied probability* pp 458–475
- Roberts GO, Rosenthal JS (2009) Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics* 18(2):349–367
- Robins G, Snijders T, Wang P, Handcock M, Pattison P (2007) Recent developments in exponential random graph (p^*) models for social networks. *Social Networks* 29(2):192–215
- Snijders TAB, Pattison PE, Robins GL, S HM (2006) New specifications for exponential random graph models. *Sociological Methodology* 36:99–153
- Tierney L (1994) Markov chains for exploring posterior distributions. *The Annals of Statistics* pp 1701–1728
- Tierney L (1998) A note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability* pp 1–9
- Tierney L, Mira A (1999) Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in medicine* 18(1718):2507–2515
- Zachary W (1977) An information flow model for conflict and fission in small groups. *Journal of Anthropological Research* 33:452–473